Sponsoring UR Data Science Capstone/Practicum Projects
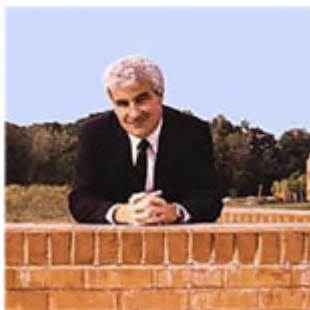
2019-Fall

Rob Rolleston, PhD

Manager, Data Science

PAYCHEX®
Payroll | Benefits | HR | Insurance

History of Paychex

1971

1983

A leading provider of integrated human capital management, providing businesses the freedom to succeed.

Payroll Services, Time and Attendance, HR Services, PEO, Hiring Services, Employee Benefits, Business Insurance, Finances and Payments, Startup Services
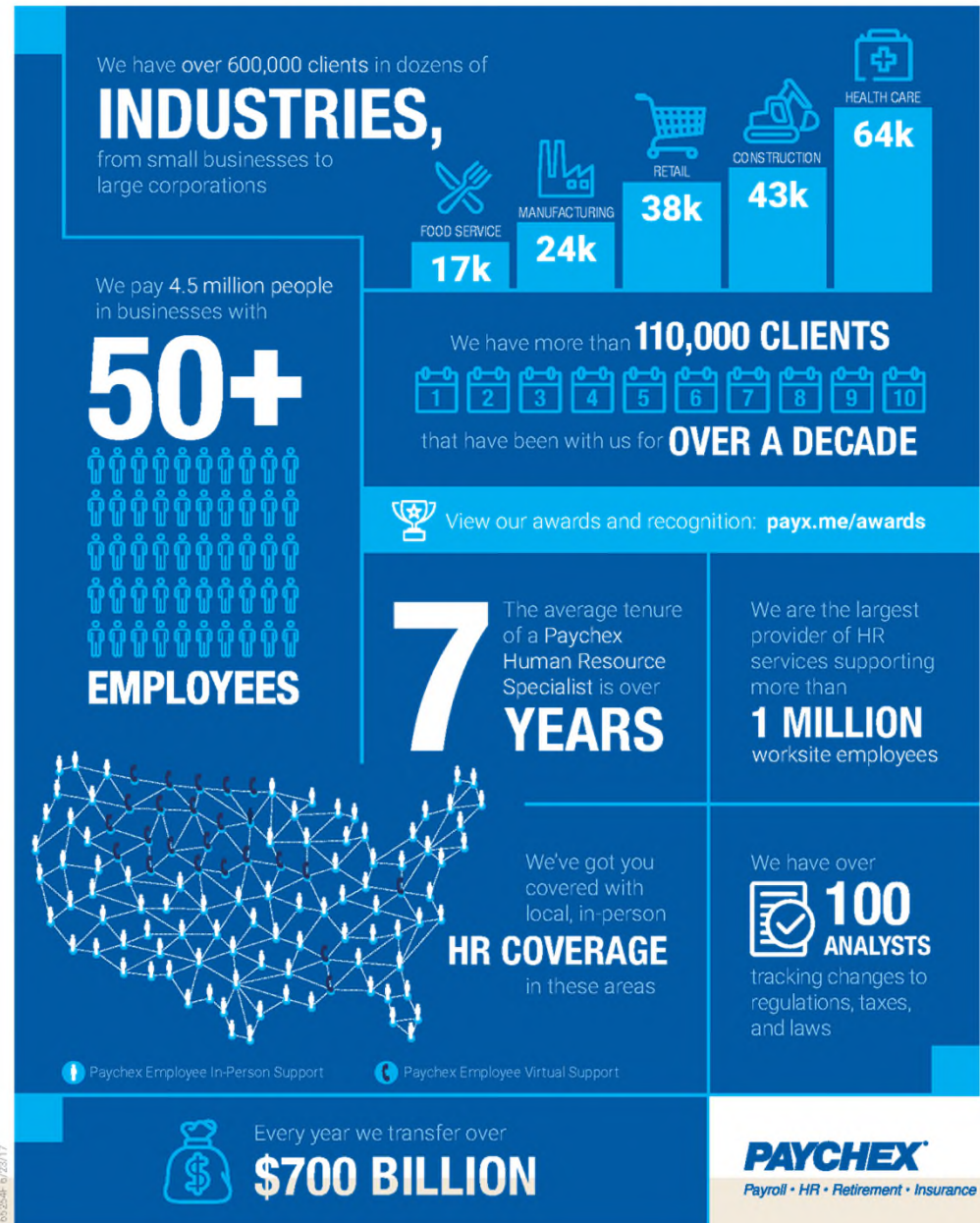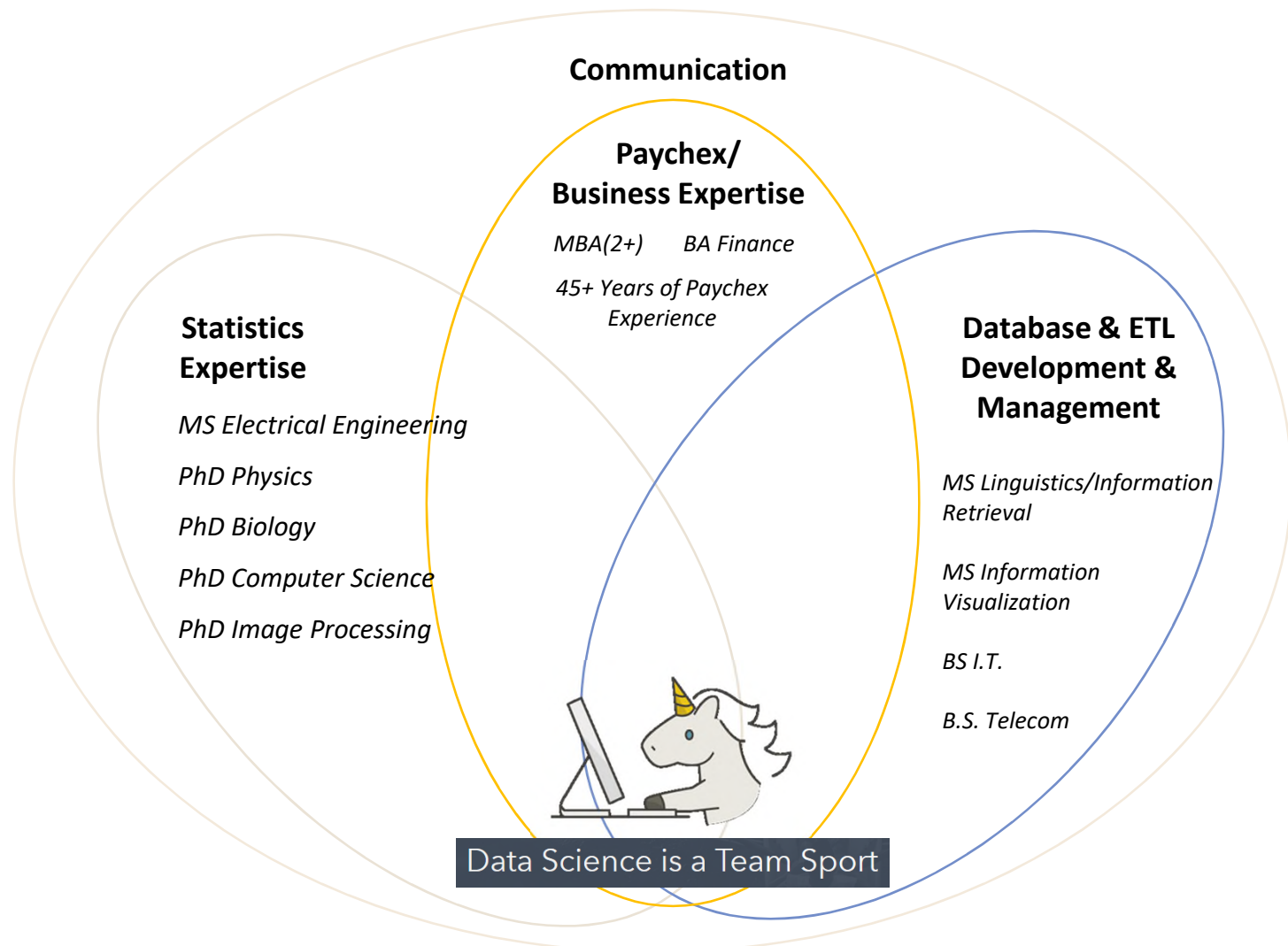
Market Cap
$30 billion

Total Revenue
$3.3 billion

Paychex Today

- Paychex is the **leading provider of** integrated payroll, retirement, and human resource **solutions for small to mid-sized businesses**.

- Paychex serves over **650,000 payroll clients** across 100 locations and pays one out of every 12 American private sector employees.
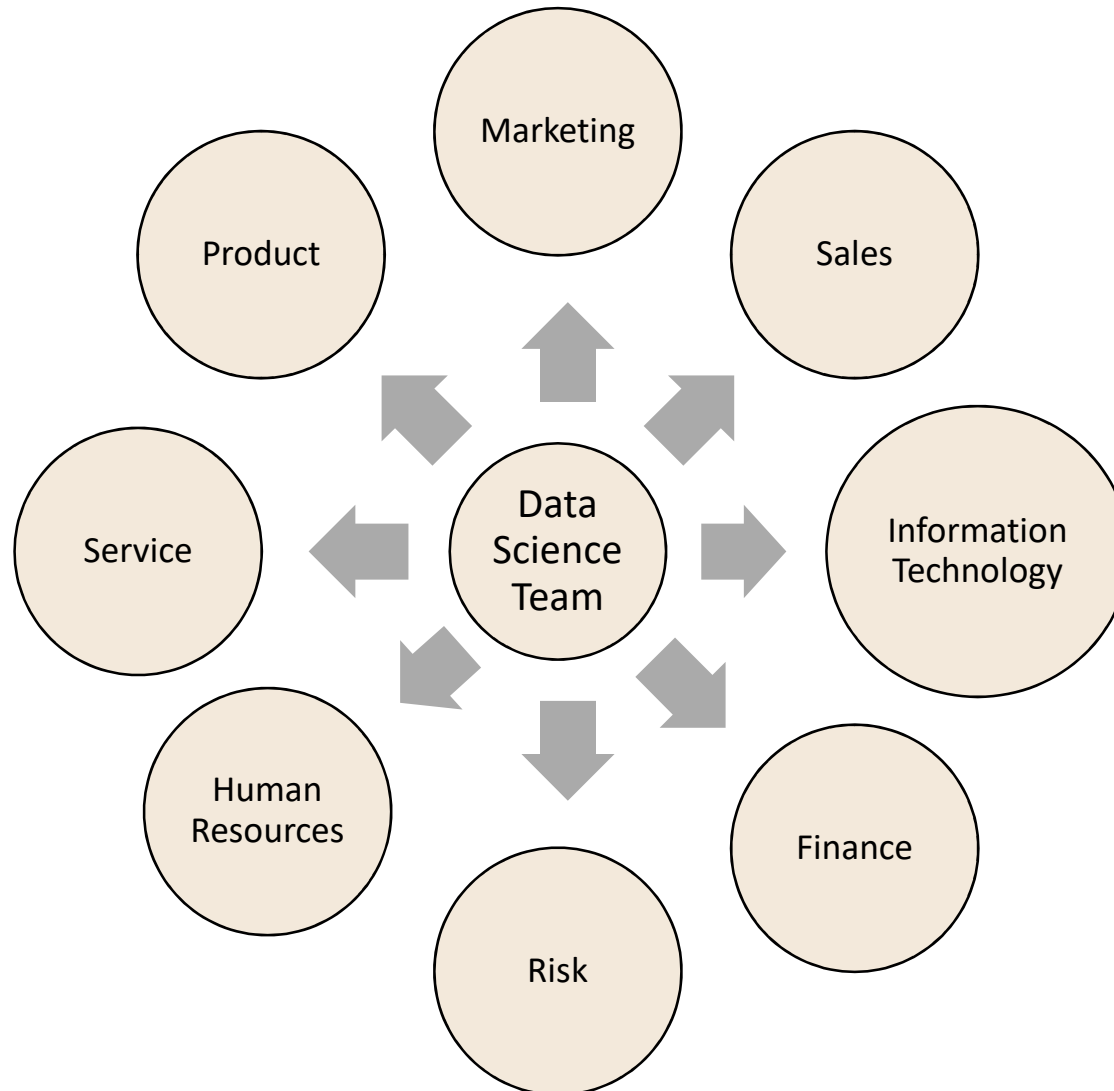
Empowering Businesses for Over 45 Years

We have over 600,000 clients in dozens of **INDUSTRIES,** from small businesses to large corporations

FOOD SERVICE **17k**

MANUFACTURING **24k**

RETAIL **38k**

CONSTRUCTION **43k**

HEALTH CARE **64k**

We pay 4.5 million people in businesses with

50+ EMPLOYEES

We have more than **110,000 CLIENTS** that have been with us for **OVER A DECADE**

1 2 3 4 5 6 7 8 9 10

View our awards and recognition: **payx.me/awards**

7 YEARS The average tenure of a Paychex Human Resource Specialist is over

We are the largest provider of HR services supporting more than **1 MILLION** worksite employees

We've got you covered with local, in-person **HR COVERAGE** in these areas

We have over **100 ANALYSTS** tracking changes to regulations, taxes, and laws

Paychex Employee In-Person Support   Paychex Employee Virtual Support

Every year we transfer over **$700 BILLION**

PAYCHEX
Payroll · HR · Retirement · Insurance

165 2E4F 6/23/17

PAYCHEX®
Payroll | Benefits | HR | Insurance

3

Data Science @ Paychex

Communication

**Paychex/
Business Expertise**

MBA(2+)   BA Finance

45+ Years of Paychex
Experience

**Statistics
Expertise**

MS Electrical Engineering

PhD Physics

PhD Biology

PhD Computer Science

PhD Image Processing

**Database & ETL
Development &
Management**

MS Linguistics/Information
Retrieval

MS Information
Visualization

BS I.T.

B.S. Telecom

Data Science is a Team Sport

Data Science @ Paychex

What is a recommendation engine?

- Data filtering tools that make use of algorithms and data to recommend the most relevant items to a particular user.
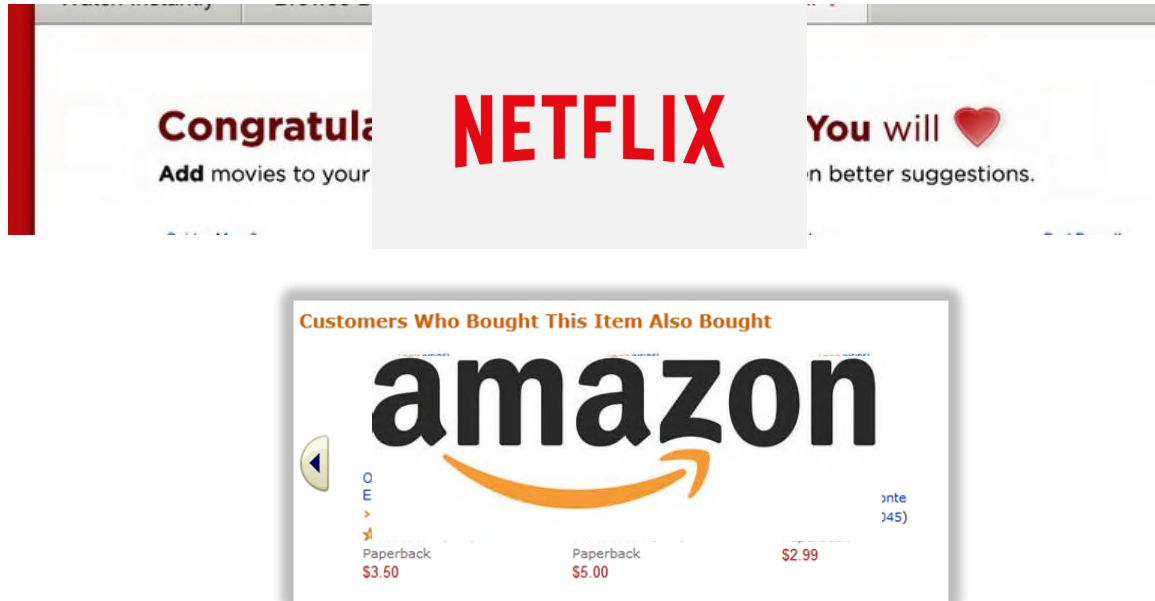
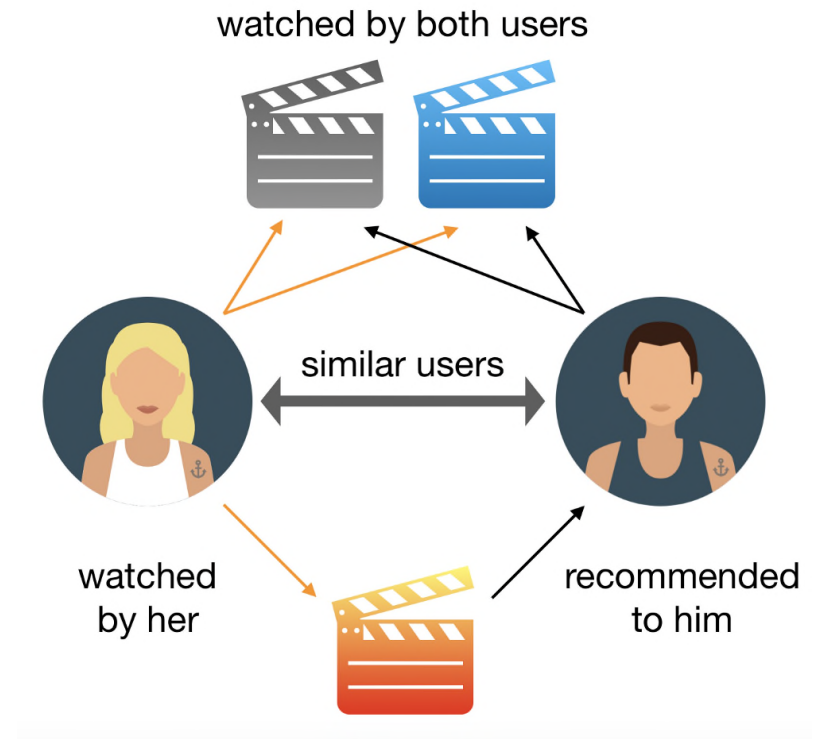# Recommendation techniques widely used

❖Content
  ➢ Based on what we like,
    ➢ Pick items with similar content to recommend us.
  ➢ Uses the content of each item for recommending purposes
    ➢ Tries to solve problems like cold-start for new users, new-item problem, etc.
  ➢ Less diversity in the recommendations,
    ➢ Will work if either the user rates things or not.

watched
by user

similar
movies

recommended
to user
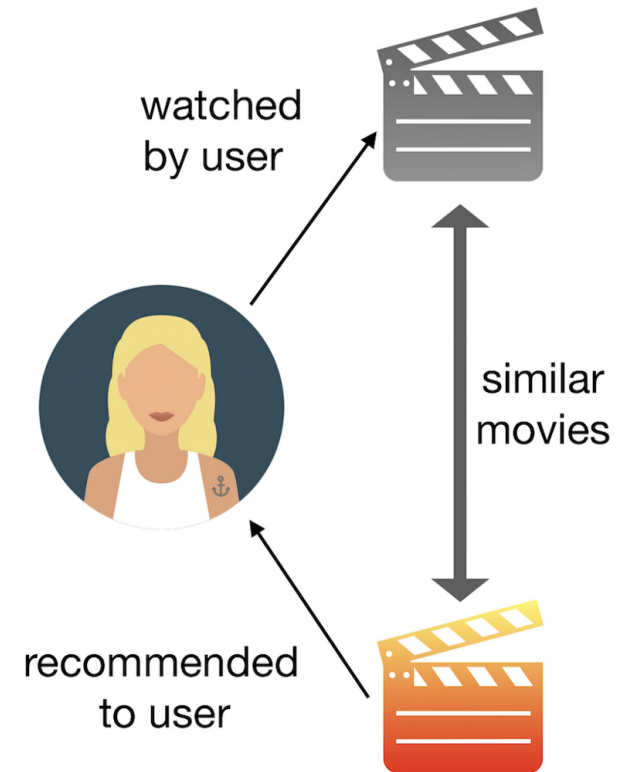
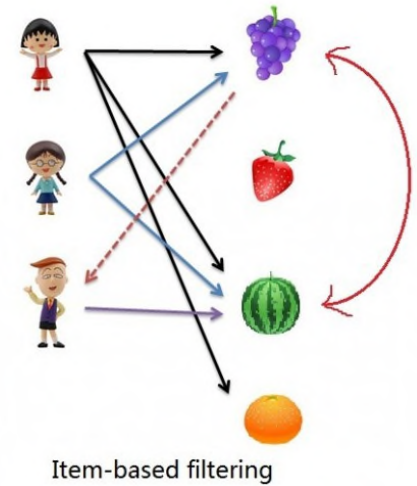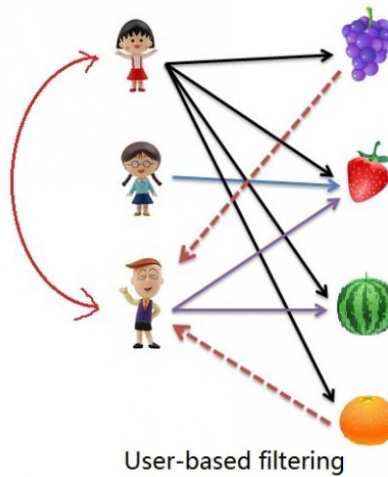# Recommendation techniques widely used

❖ Collaborative

➢ Does not require any information about the item or the users

  ➢ Recommendations based on users' past behaviors

➢ Based on users' ratings,

  ➢ Recommends items that we don't have, but users similar to us do have

➢ Need to have some form of rating for this algorithm to work

watched by both users

similar users

watched by her

recommended to him

Recommendation techniques widely used

❖ **Hybrid: Collaborative Filtering Types**
  ➤ User-based recommendation:
    ➤ Find the similar users and recommend items.
  ➤ Item-based recommendation:
    ➤ The similarity of items is calculated and items are recommended.

User-based filtering

Item-based filtering

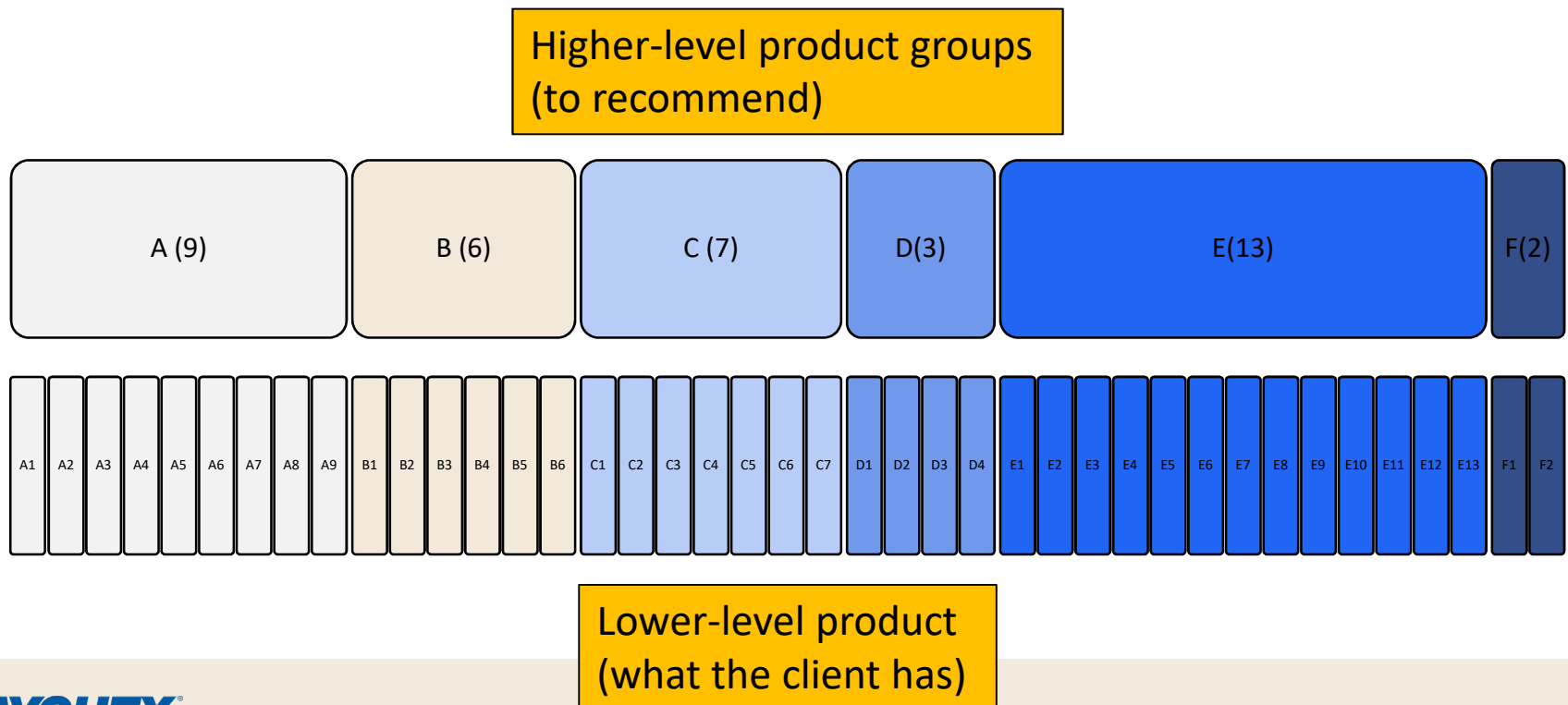# Recommendation techniques widely used

❖ **Association Rules**

➢ "Market Basket Analysis"

➢ Discovering interesting relations between items in large databases

➢ Sets of items are often associated with one another
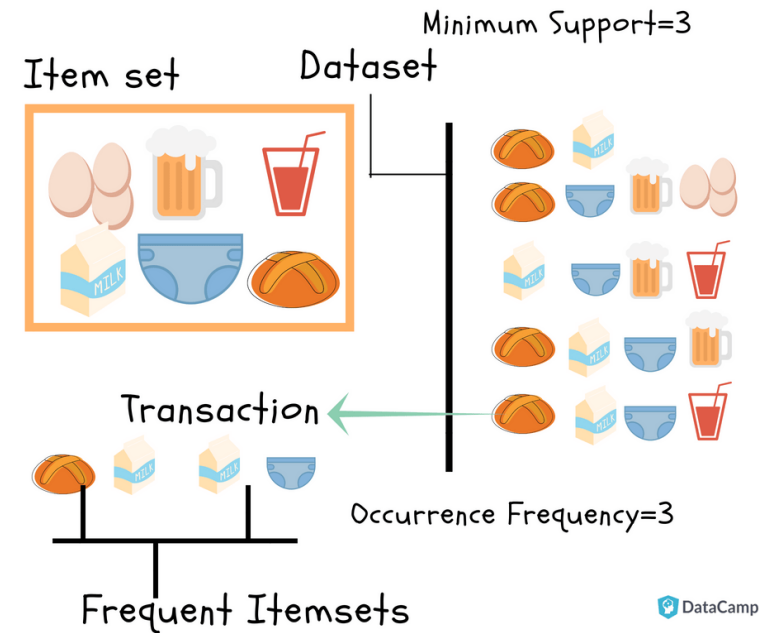
Recommending a new product to any client who calls in, which will eventually turn into a sale

- PAYX has a finite amount of products which it offers to all of its clients
- Products are clustered into 6 high level product groups
- We want to recommend a specific <u>high-level</u> product categories to clients who call us

Higher-level product groups (to recommend)

A (9) | B (6) | C (7) | D(3) | E(13) | F(2)

A1 A2 A3 A4 A5 A6 A7 A8 A9 B1 B2 B3 B4 B5 B6 C1 C2 C3 C4 C5 C6 C7 D1 D2 D3 D4 E1 E2 E3 E4 E5 E6 E7 E8 E9 E10 E11 E12 E13 F1 F2

Lower-level product (what the client has)

PAYCHEX®
Payroll | Benefits | HR | Insurance

Data set overview

- One File: ~ 50MB
  - 44 columns
    - See below for column description
  - 382,525 rows
    - One row per client

ID	A1	A2	A3	A4	A5	D1	B1	D2	A6	C1	C2	D3	C3	C5	C6	B2	A7	C7	A8	E1	E2	E3	E4	E5	E6	E7	D4	E8	A9	F1	F2	B3	B4	B5	E9	E10	E11	E12	E13	B6	Rep Level	Size	Industry
1	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Level 4	Size 3	Industry 10
2	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	Level 5	Size 4	Industry 6
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Level 5	Size 4	Industry 10
4	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Level 4	Size 2	Industry 12
5	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Level 4	Size 3	Industry 10
6	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	1	0	0	Level 4	Size 4	Industry 10

Client ID

Low level products.
1 – Has the product
0 – Doesn't have
the product

Categorical info
about the client.
(more info on the
next slide)

Categorical Variables = Client Description

- Rep Level – Which type of rep will make recommendation to the client?
  - 6 levels – Level 1, Level 2, Level 3, Level 4, Level 5, Unknown
  - These levels are not ordered, meaning Level 1 Rep is not necessarily better/worse than Level 5 Rep.
- Size – What is the size of the client (in terms of employees the client has)?
  - 5 Sizes – Size 1, Size 2, Size 3, Size 4, Size 5
  - These are ordered Size 1 < Size 2 < Size 3 < Size 4 < Size 5
- Industry – Which type of industry does the client belong to?
  - 13 distinct industry types.
  - Industry 1, Industry 2, …….. , Industry 12, Industry 13

Expected Outcomes

- Recommendation engine algorithm(s)
  - Priority list of group(s) | product(s) to recommend to each client
  - To segment the 400K clients or not….

- Performance Metrics
  - How do we compare the performance of different algorithms?
  - Caveat: Traditional metrics like AUC do not work

- Documents
  - Regular meetings & reports
    - We're here to collaborate and advise
  - Code
    - R preferred,   but we suspect you all love Python
  - Final Presentation

Questions ?
Discussion…

PAYCHEX®
Payroll | Benefits | HR | Insurance